Microsoft

Analytics in a day

Cloud analytics in the age of self-service and data science

Microsoft

Steen Kjøng Kristensen

Senior Consultant, Business Intelligence
Accobat A/S

skp@accobat.com
https://www.linkedin.com/in/spauls/

Agenda

Times are approximate and will be fluid with the class

Workshop Instruction & Discussion

9:00 – 9:30 Harness the power of analytics
9:30 – 10:30 Experience a new class of analytics with Azure Synapse
10:30 – 10:45 Get started with your first analytics project

Break for 15 min

Hands-on lab

11:00 – 12:30 Hands-on lab using Azure Synapse and Power BI

Break for 30 min

Hands-on lab

13:00 – 15:30 Power BI Extension Module

Microsoft

Harness the power of analytics

Section 1
All businesses are data businesses

Section 2
The cloud for modern analytics

Section 3
The paradox of analytics

Section 4
The analytics continuum

Microsoft

Section 1

All businesses are data businesses

Today, all businesses are data businesses

Data is the lifeblood of modern work

Data makes business possible

Data is crucial to businesses of all sizes

Securing data is mandatory

All data businesses need to be analytic businesses

Without analytics data is a cost center, not a resource

Businesses need to “know thyself”

Workers are empowered with intelligent insights

Augmented intelligence sees patterns we cannot

Analytic businesses need to evolve data science

Every business has opportunities to make analytics faster, easier, and more insightful

Data science empowers us to go farther

AI/ML creates smarter organizations

Businesses need scalable analytics

The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States

Volume
SCALE OF DATA

40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005

It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day

6 BILLION PEOPLE
have cell phones

WORLD POPULATION: 7 BILLION

Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored

Variety
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]

30 BILLION PIECES OF CONTENT
are shared on Facebook every month

By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS

4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month

400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Velocity
ANALYSIS OF STREAMING DATA

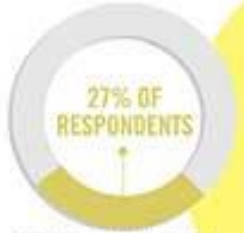
The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth

Veracity
UNCERTAINTY OF DATA

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



in one survey were unsure of how much of their data was inaccurate

Poor data quality costs the US economy around
$3.1 TRILLION A YEAR

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, GAS

Microsoft

Section 2

The cloud for modern analytics

Modern businesses succeed in the cloud

The cloud is the default environment for new technology initiatives

Applications are increasingly cloud-native

Collaboration platforms and file-sharing are transforming

Hybrid businesses are replacing on-premises data management

A cloud analytics platform is an economic breakthrough

Price, performance, and agility

High infrastructure costs are avoided with cloud-based analytics platforms

Iterating designs and handling diverse data becomes more efficient

Elastic scaling makes world-class performance available to all users

A cloud analytics platform is the hub for all data models

Structured, unstructured, and streaming data integrated in a single, scalable, environment

Workloads can be stored, managed and provisioned together

No need to juggle processing schedules

Visualization and data science tools empower enterprise-wide insights

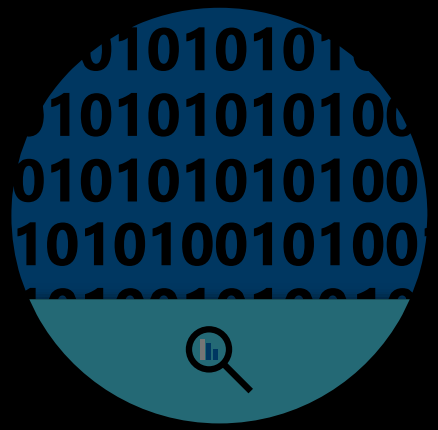
Microsoft

Section 3

The paradox of analytics

While data grows 400%, less than 30% gets analyzed

2020



4x

2025

44ZB Data collected worldwide

175ZB Data collected worldwide

More analytics solutions lead to more silos

Data

Structured
Unstructured
Streaming
Big data
Cloud
On-premises
IoT/edge

Technologies

Map Reduce
Open source projects
Data mart
Data warehouse
Data lake
Spreadsheets
RDMS
Data visualization tools
ML models
AI services

Skills

SQL
Python
Java
R
Industry Schemas
Data modeling
Data cleaning
Data integrations

This leads to the paradox of analytics

Each new technology creates another siloed operation

Big data

Data integration

Machine learning

Business intelligence

Data governance

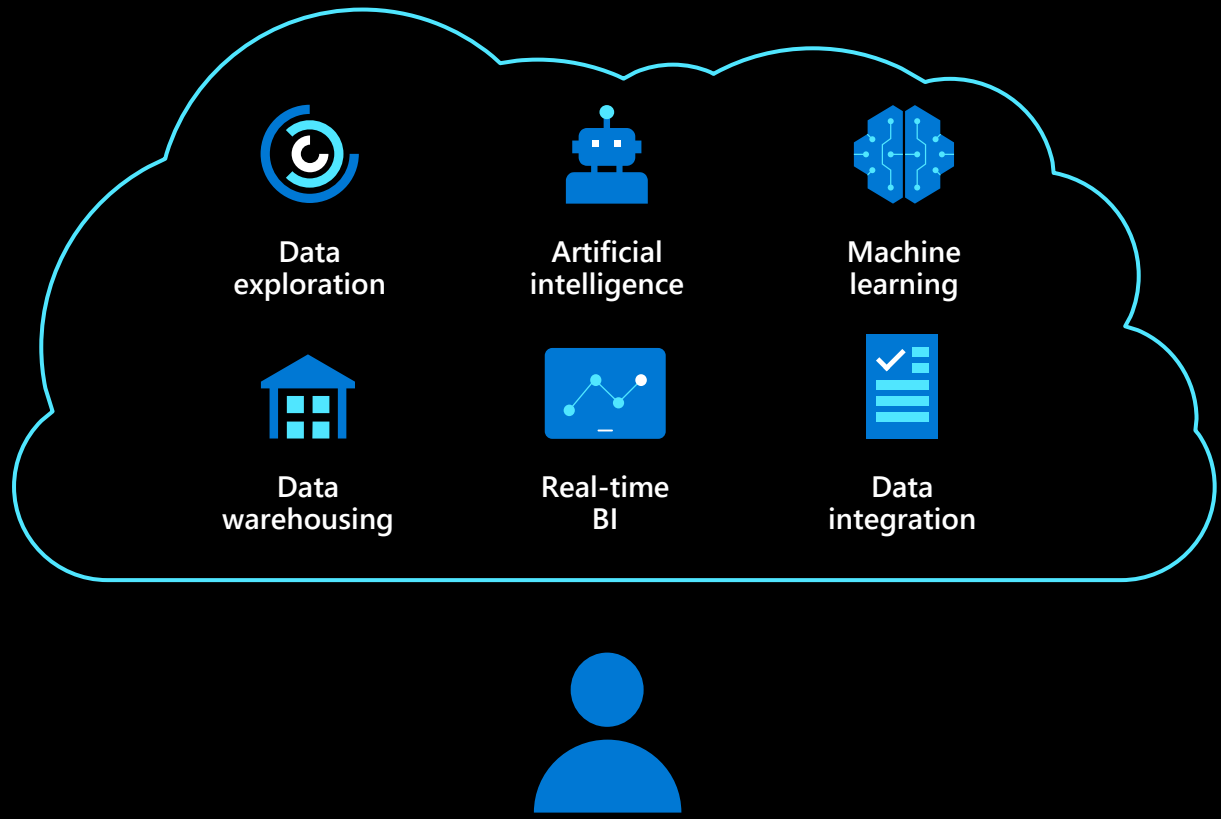
Security

Microsoft

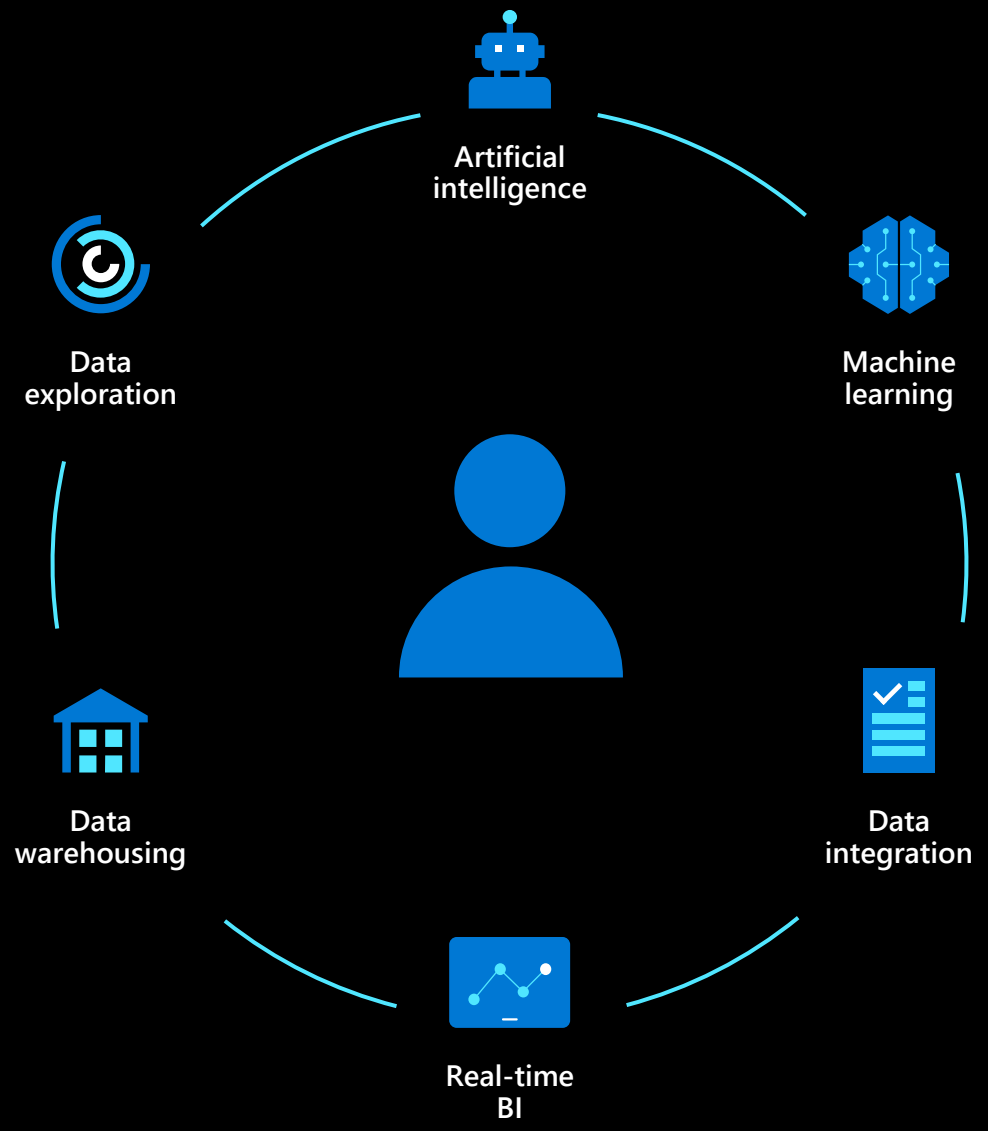
Section 4

The analytics continuum

Data professionals shouldn’t have to continually learn new skills to deliver insights



Analytics should seamlessly be part of the way users work—rather than a factor for creating more siloes

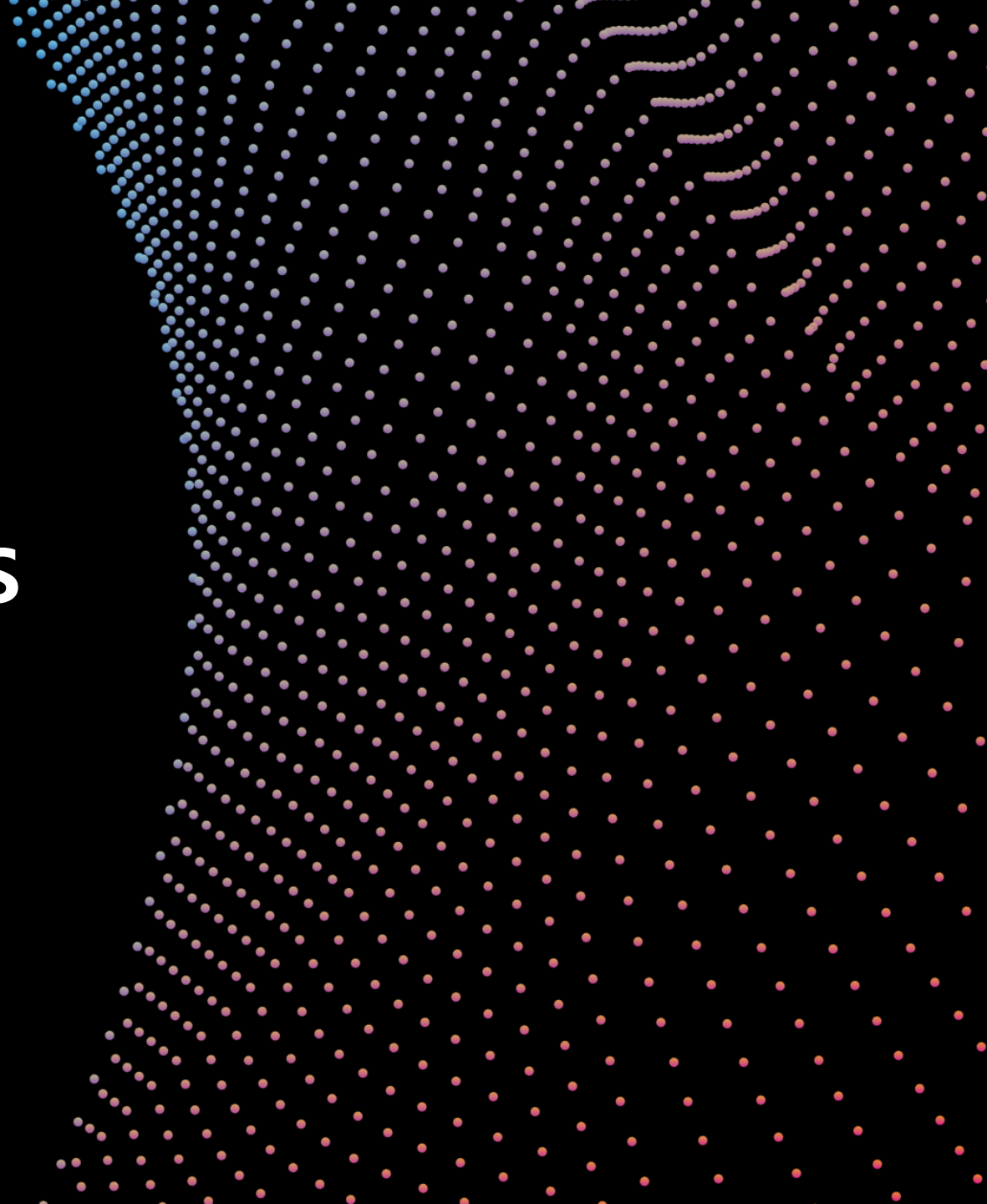


Azure Synapse delivers the full continuum of analytics in a single service

Microsoft

Azure Synapse Analytics

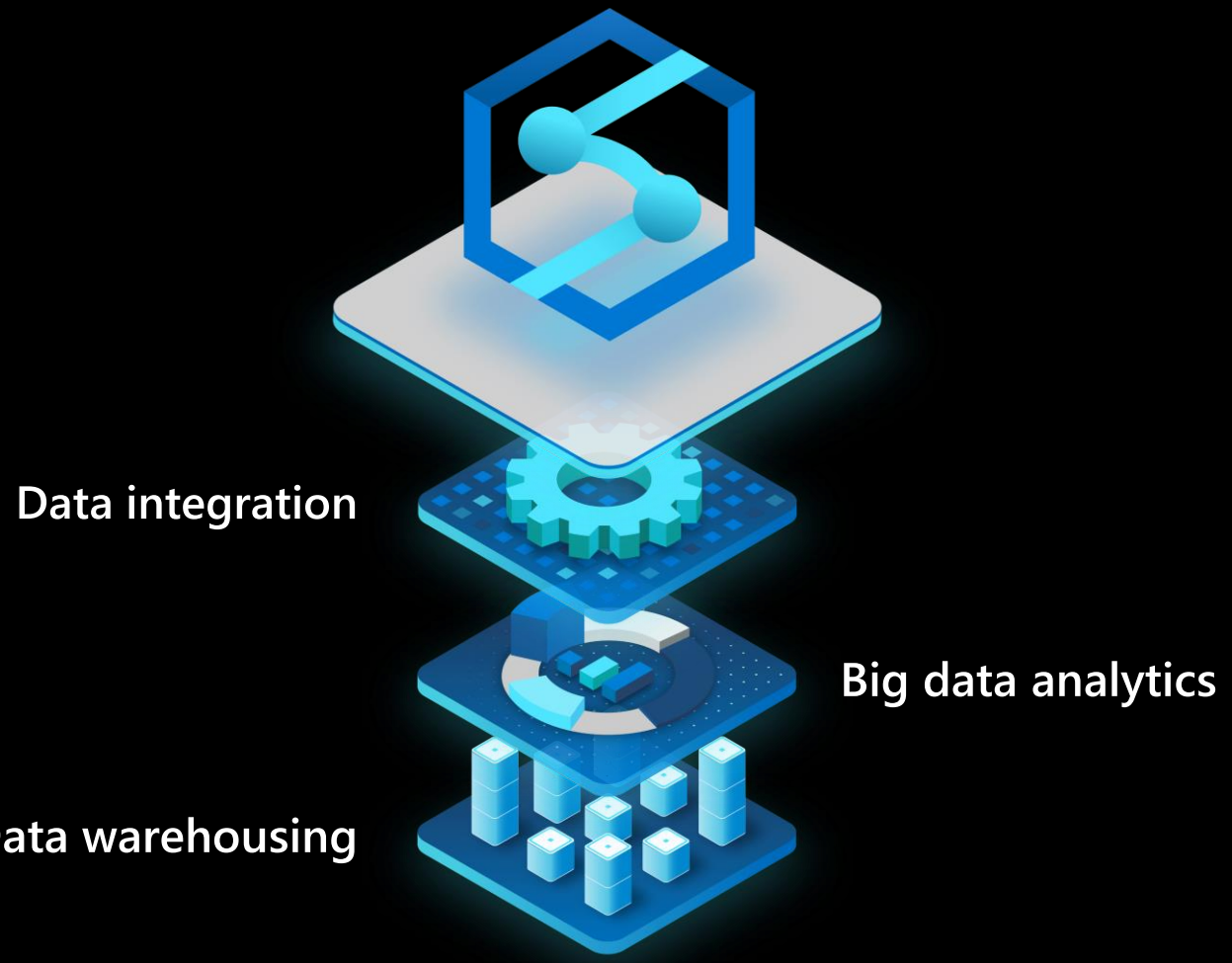
Limitless analytics service with unmatched time to insight



Azure Synapse Analytics

The first unified, cloud native platform for converged analytics

Azure Synapse is the only unified platform for analytics, blending big data, data warehousing, and data integration into a single cloud native service for end-to-end analytics at cloud scale.



Azure Synapse Analytics

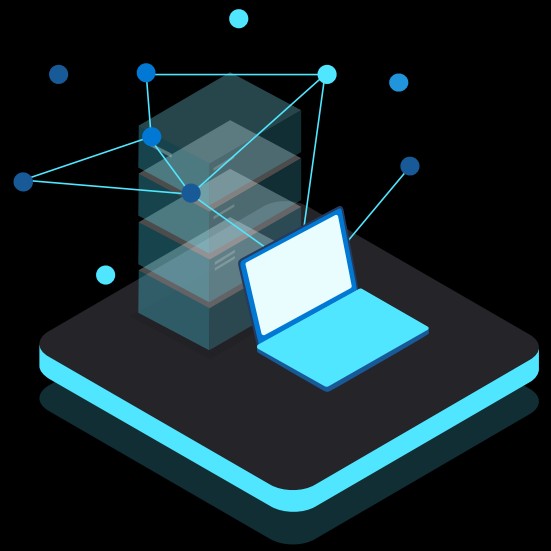
Powered by a new cloud native distributed SQL engine

Serverless + dedicated SQL

Flexible consumption models

Serverless pay-per-query ideal for ad-hoc data lake exploration and transformation

Dedicated clusters optimized mission-critical data warehouse workloads



Serverless

Dedicated

Microsoft

Let’s dive deeper

- Comprehensive security and compliance
- Streamlined data integration
- Flexible data warehousing
- Real-time operational analytics
- Integrated machine learning
- Power BI + Azure Synapse

Security and compliance

Unified governance controls

Complete data protection

Best-in-class security

Customer & System Managed Keys

All data encrypted by default

Up to 3x levels of data encryption at rest

Democratize data at scale with fine-grained ACL

Proactive protection

Comprehensive Compliance

Category	Feature	
Data Protection	Data in transit	✓
	Data encryption at rest	✓
	Data discovery and classification	✓
Access Control	Object level security (tables/views)	✓
	Row level security	✓
	Column level security	✓
	Dynamic data masking	✓
	Column level encryption	✓
Authentication	SQL login	✓
	Azure active directory	✓
	Multi-factor authentication	✓
Network Security	Managed virtual network	✓
	Custom virtual network	✓
	Firewall	✓
	Azure ExpressRoute	✓
	Azure Private Link	✓
Threat protection	Threat detection	✓
	Auditing	✓
	Vulnerability assessment	✓
Isolation	Dedicated metadata store	✓
	Hosted in customer tenant	✓

Managed virtual networks

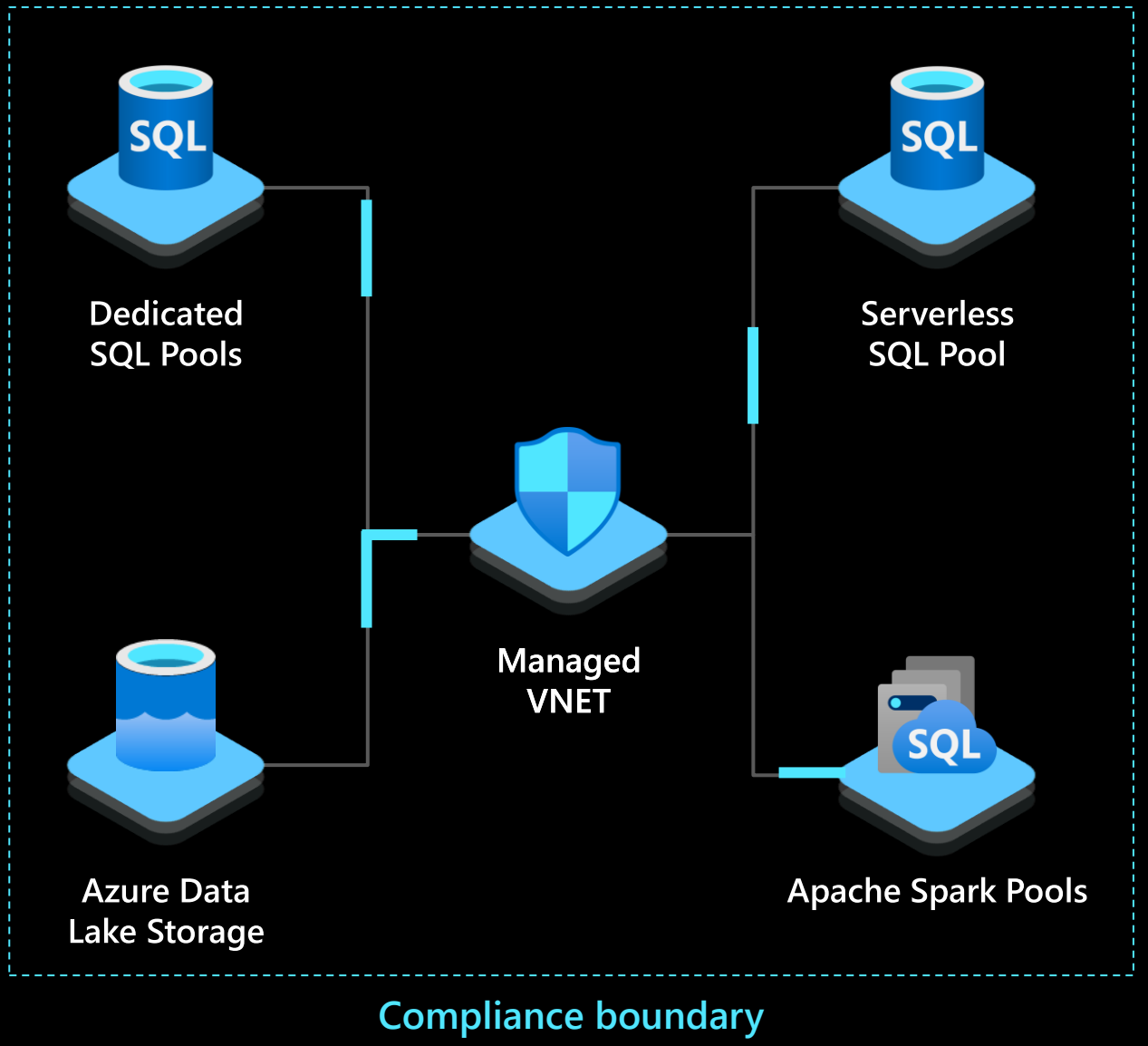
Eliminate network maintenance

One-click enables automated management of virtual networks between cluster endpoints

Synapse resources only ever interop with private endpoints

No management of subnets or IP Ranges

Prevents data exfiltration



Data integration

Code-free hybrid data integration

Hybrid data integration

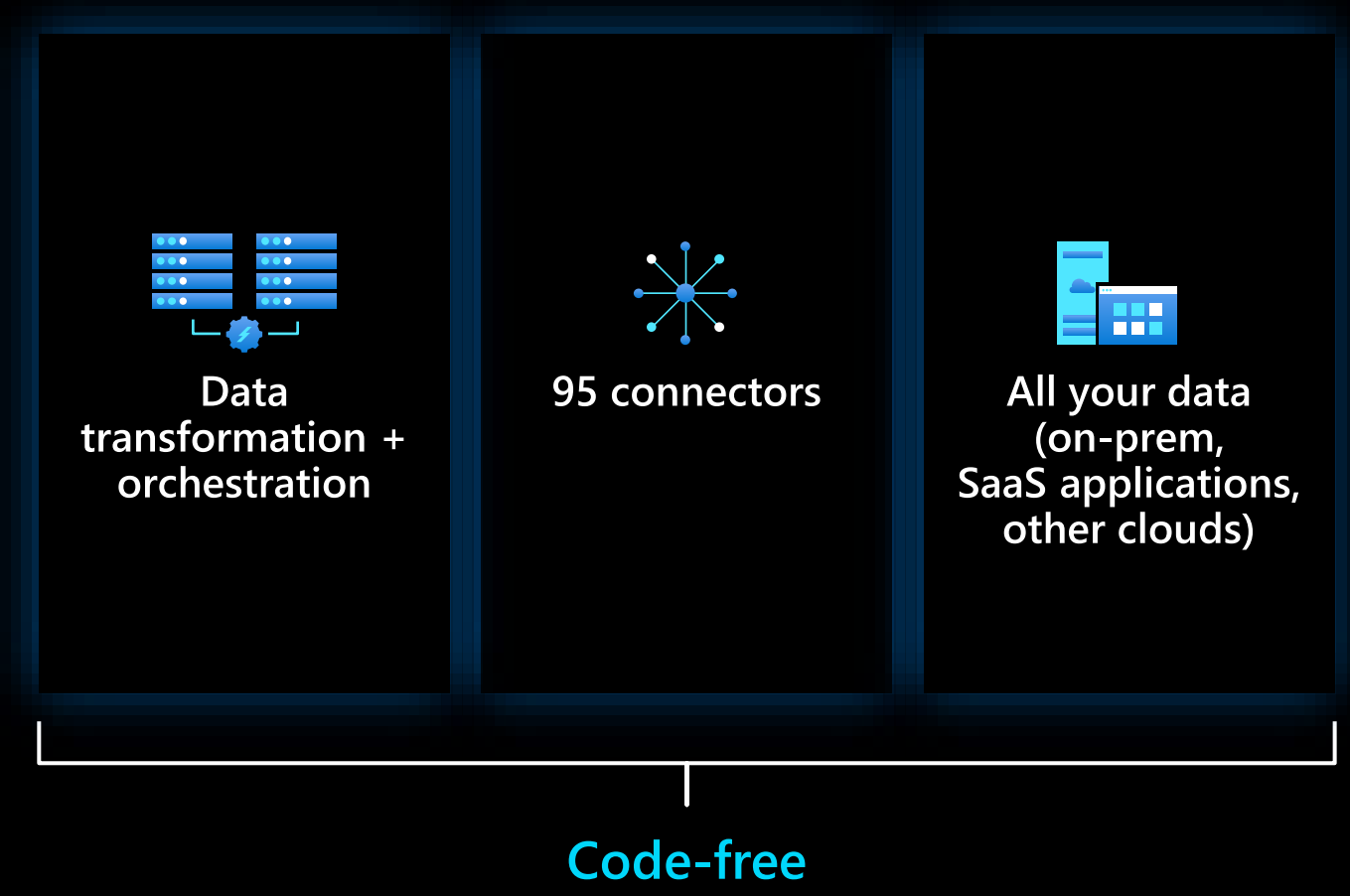
Cloud native ETL/ELT

95+ connectors available

Secure connectivity to on-premise data sources, other clouds, and SaaS applications

Code-first and low/no code design interfaces

Schedule and Event based triggering



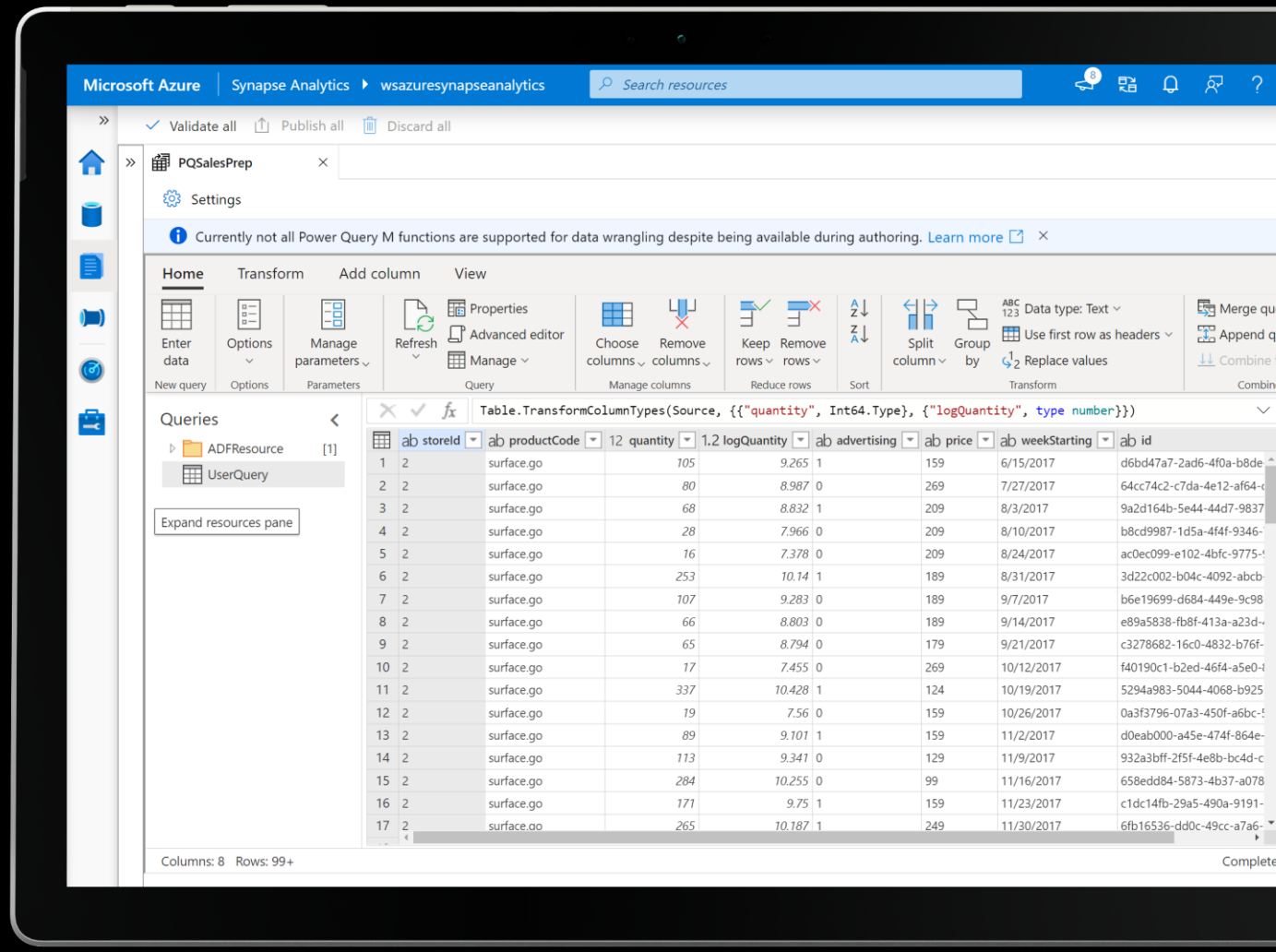
Code-free data wrangling

No/low-code data transformation

Excel-like interface is familiar to users

Transform data to desired shape completely visually

Operationalize into pipelines



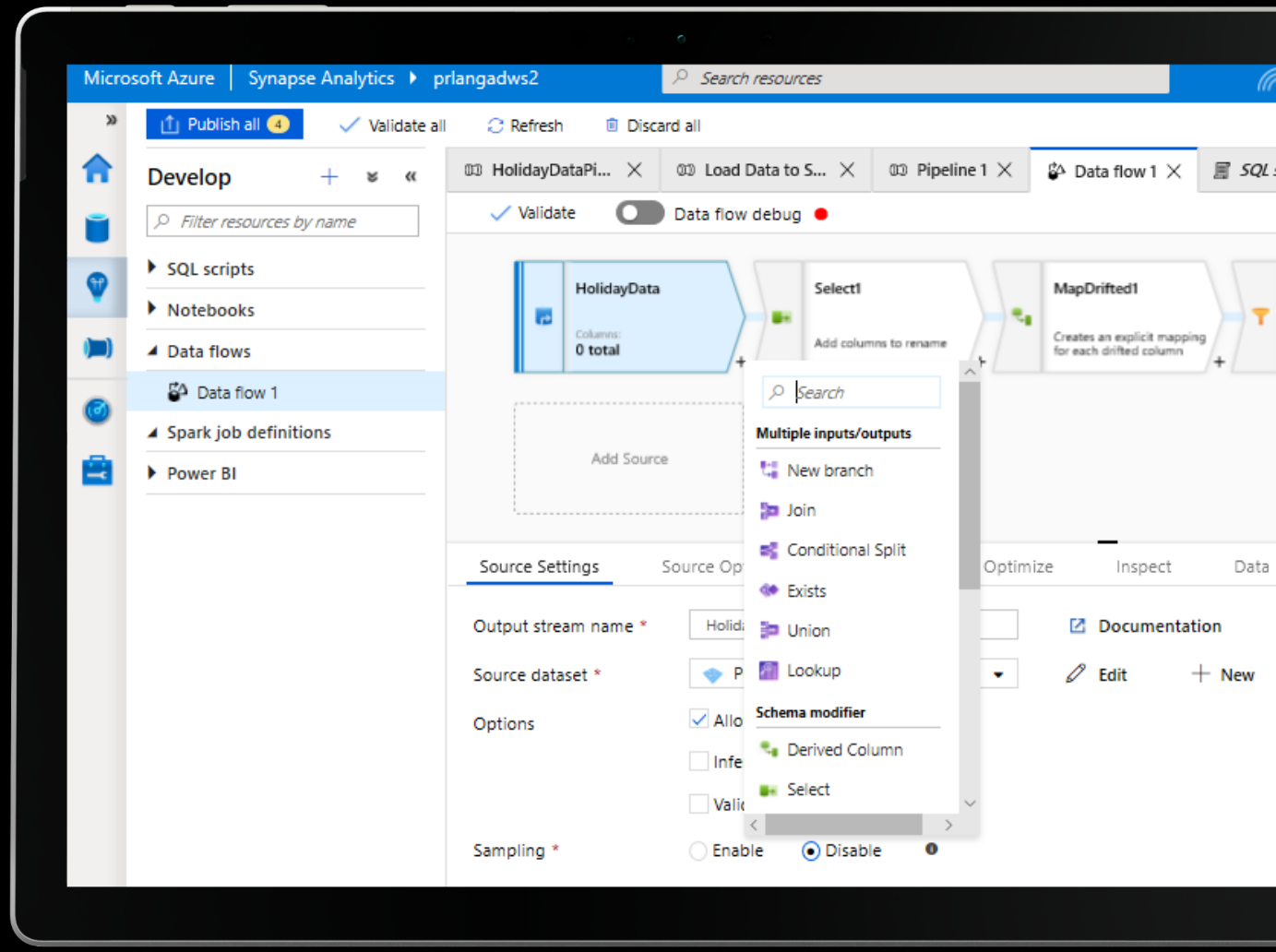
Hybrid data integration

Real-time operational analytics

No data integration pipelines required

Near-zero impact on operational systems

Latency <90s at 99th percentile



Data warehouse

Scalable and secure analytics platform for SQL workloads

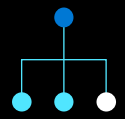
Suitable for any scale

Fully-managed elastic platform

Elastic compute that can be easily optimized to different classes of workload

All features available in a single tier

Infinite cost effective PAYG storage



Department

Gigabytes of data

10’s of users

Weekday usage

Organization

Terabytes of data

100’s of users

24/7 usage

Enterprise

Petabytes of data

1000’s of users

Tier-0 Availability

Synapse Studio

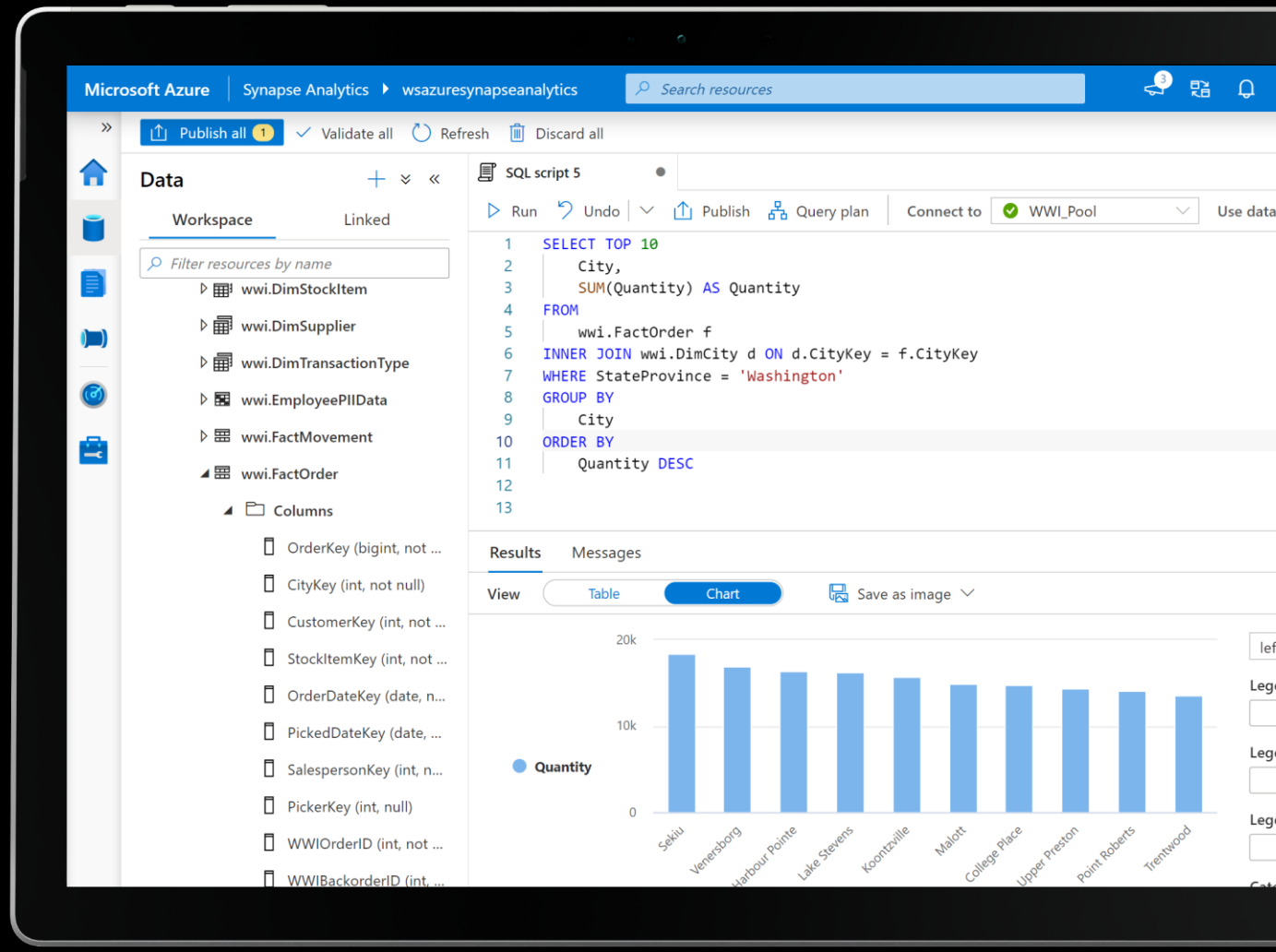
SQL Editor

Automatic code completion (Intellisense)

Script collaboration within the Workspace

Built-in visualizations

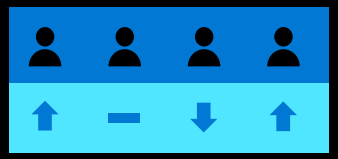
Easily switch between clusters



Workload management

Scale in

Workload Importance



Data Warehouse

Benefits:

- Predictable cost
- Bias to high-value workloads within fixed/predictable budget
- Enables customers to easily deprioritize queries which don't need to be run immediately
- Built-in starvation prevention

Workload Isolation

Data Warehouse

Benefits:

- Predictable cost
- Efficient for unpredictable workloads since compute can overcommit
- No cache eviction for scaling (no performance cliff at restart)
- Single connection string for isolation (unlike Snowflake virtual warehouses)

Scale out

Elastic Cluster (Scale Up)

Data Warehouse

Benefits:

- Incremental add compute
- Increase large query performance
- Single cache for heterogeneous workloads
- Single endpoint

Azure Synapse supports a more diverse set of workload management tools through workload importance, intra-cluster isolation, and elastic clusters.

Machine learning

Empower everyone with predictive insights

Machine learning

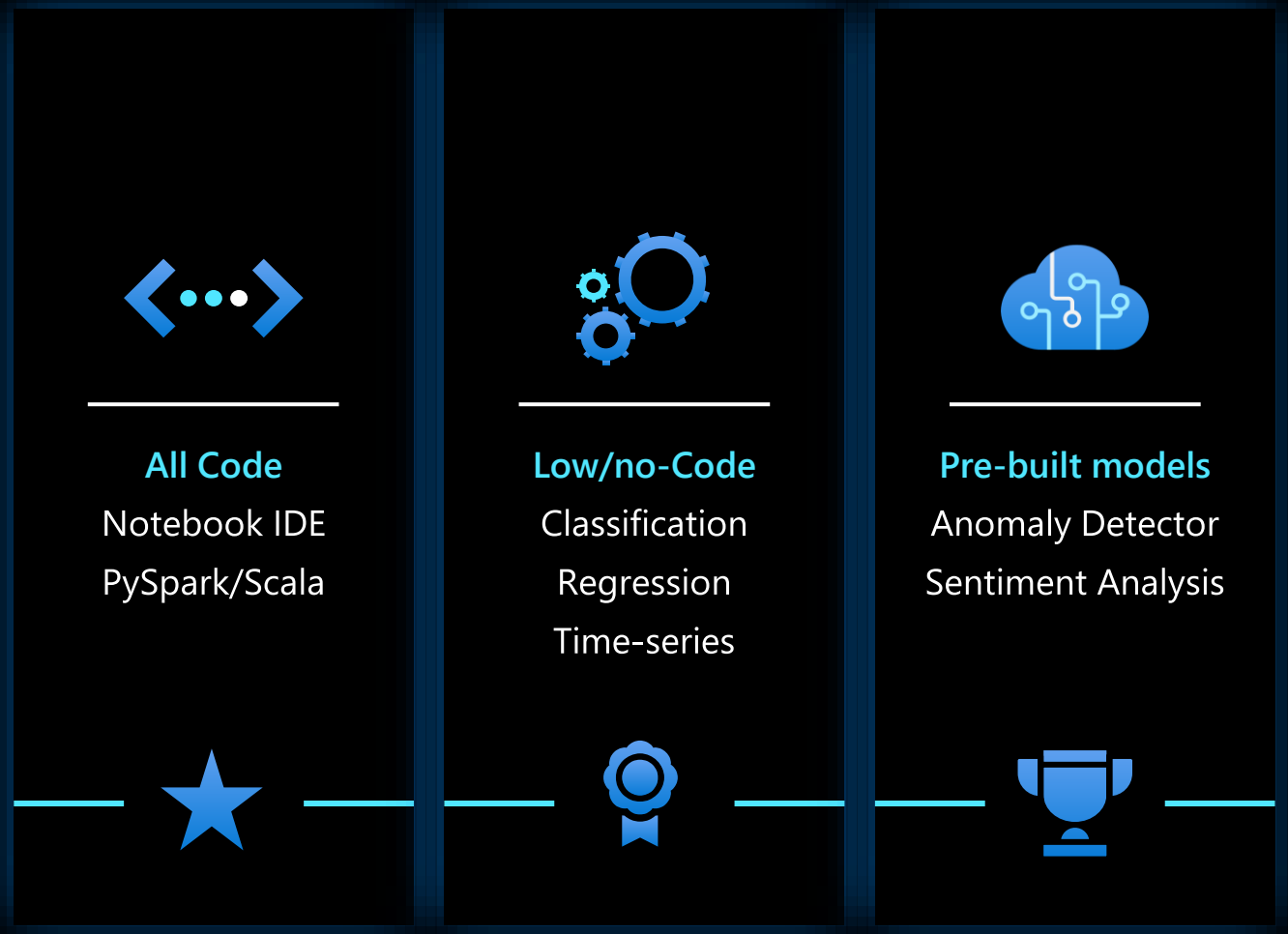
Democratize predictive power

Synapse makes predictive analytics accessible to all

Notebooks provides a code authoring experience for complex predictive models

Automatic ML graphical interface provides a no-code experience for creating ML models

Native integration with Azure Cognitive Search provides access to pre-built models



Notebook IDE code authoring

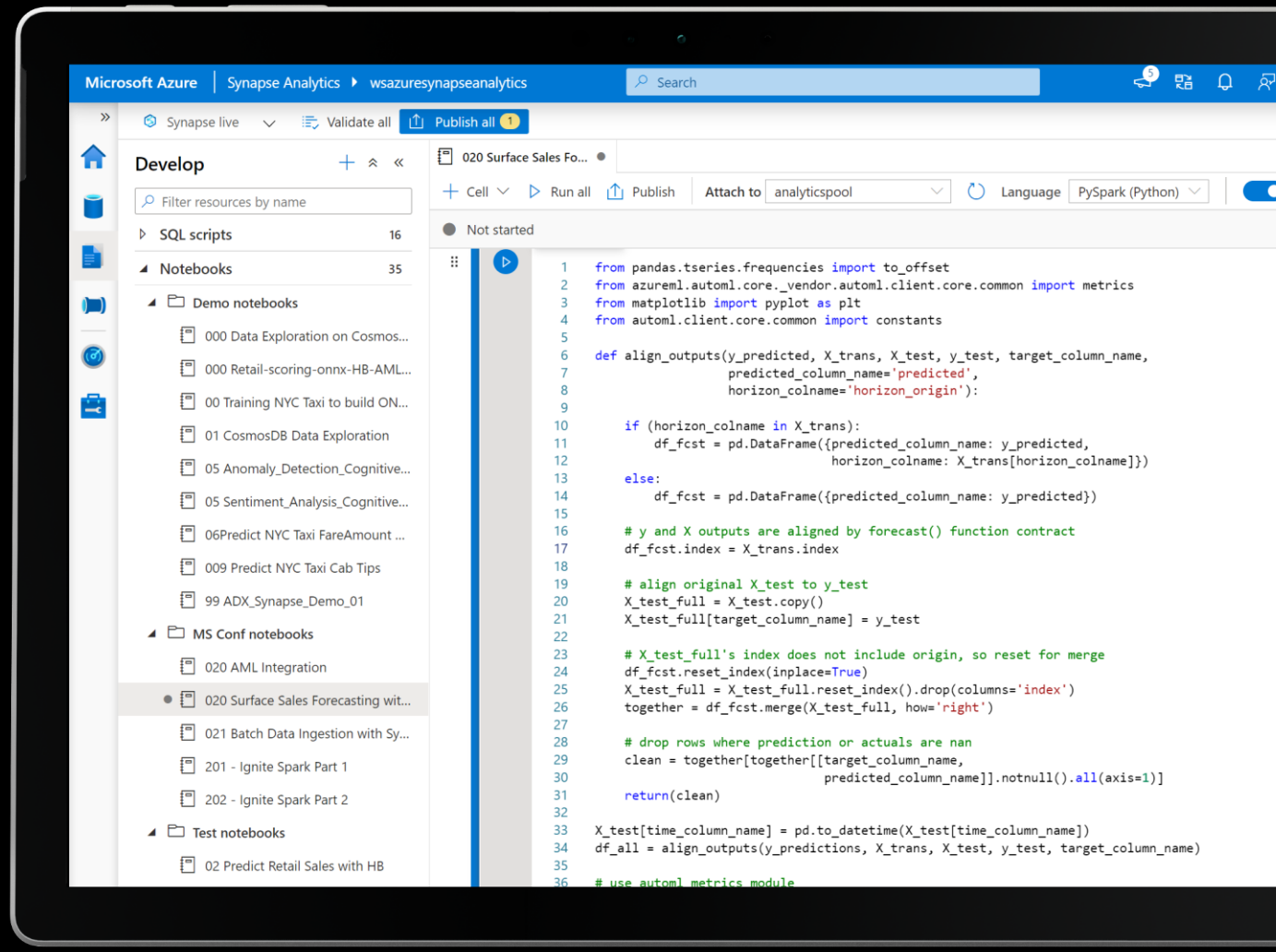
Code-first ML model development

PySpark, Scala, and C# languages supported

Automatic code completion (Intellisense)

Author multiple languages in a single notebook

Analyze data from the data warehouse, data lake, and real-time operational data from one place

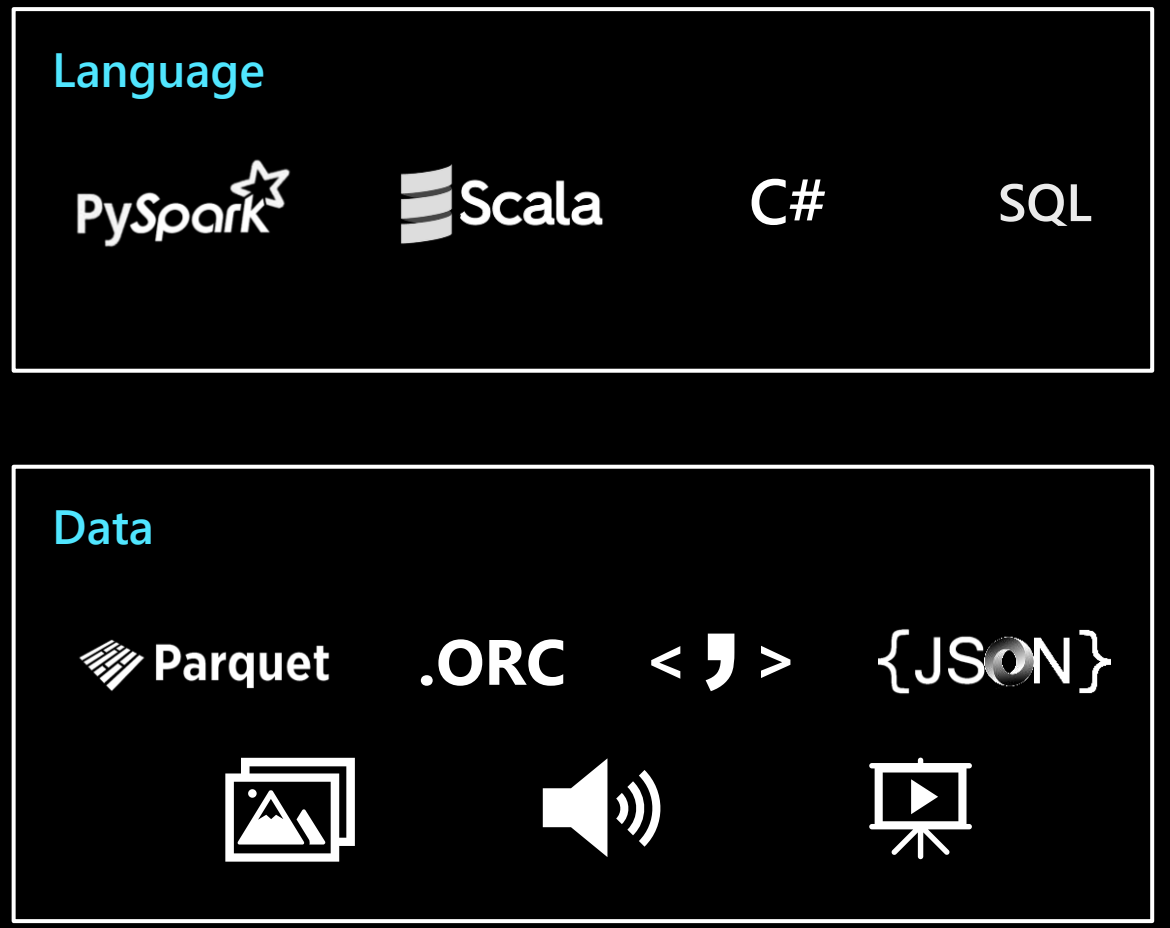


Open ecosystem with support for industry standards

Data + Languages

Languages such as SQL, PySpark, Scala and C# in support of data science and data warehouse workloads

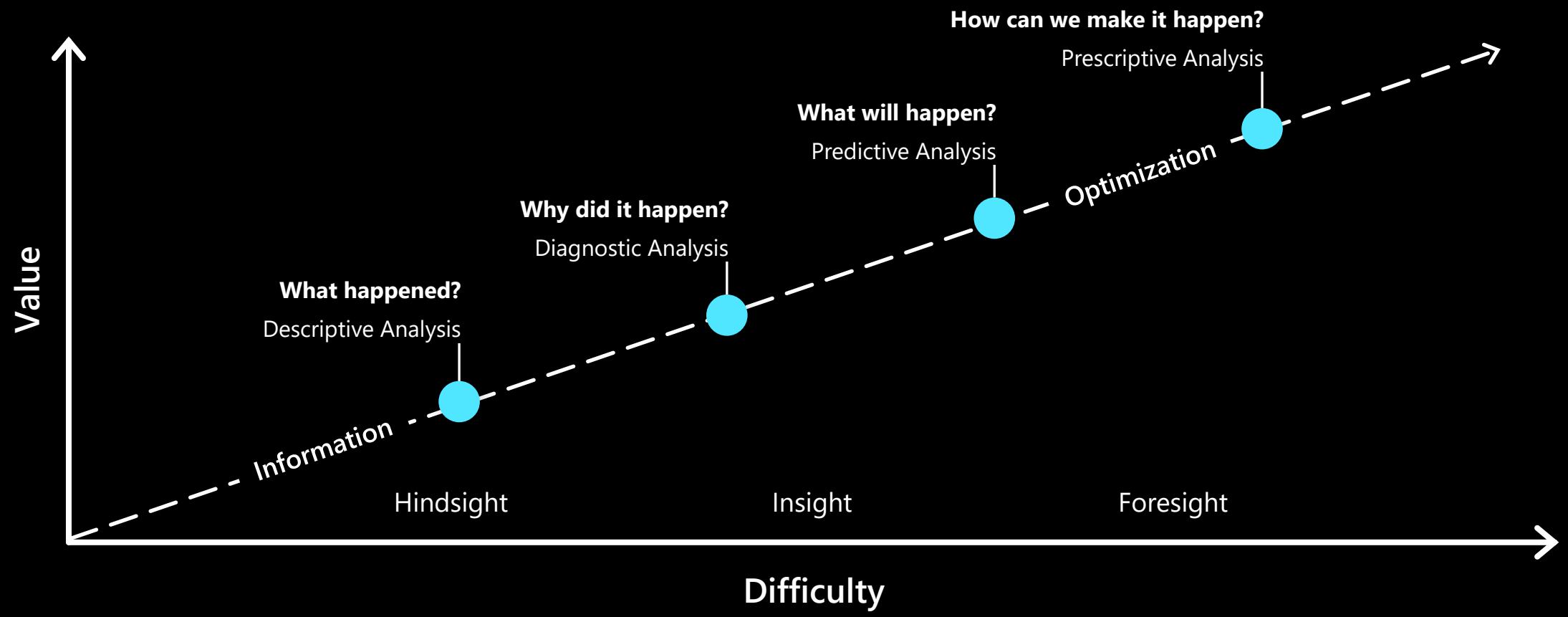
The data lake supports and unlimited set of file formats including Parquet, ORC and Json as well as audio, image, and video formats



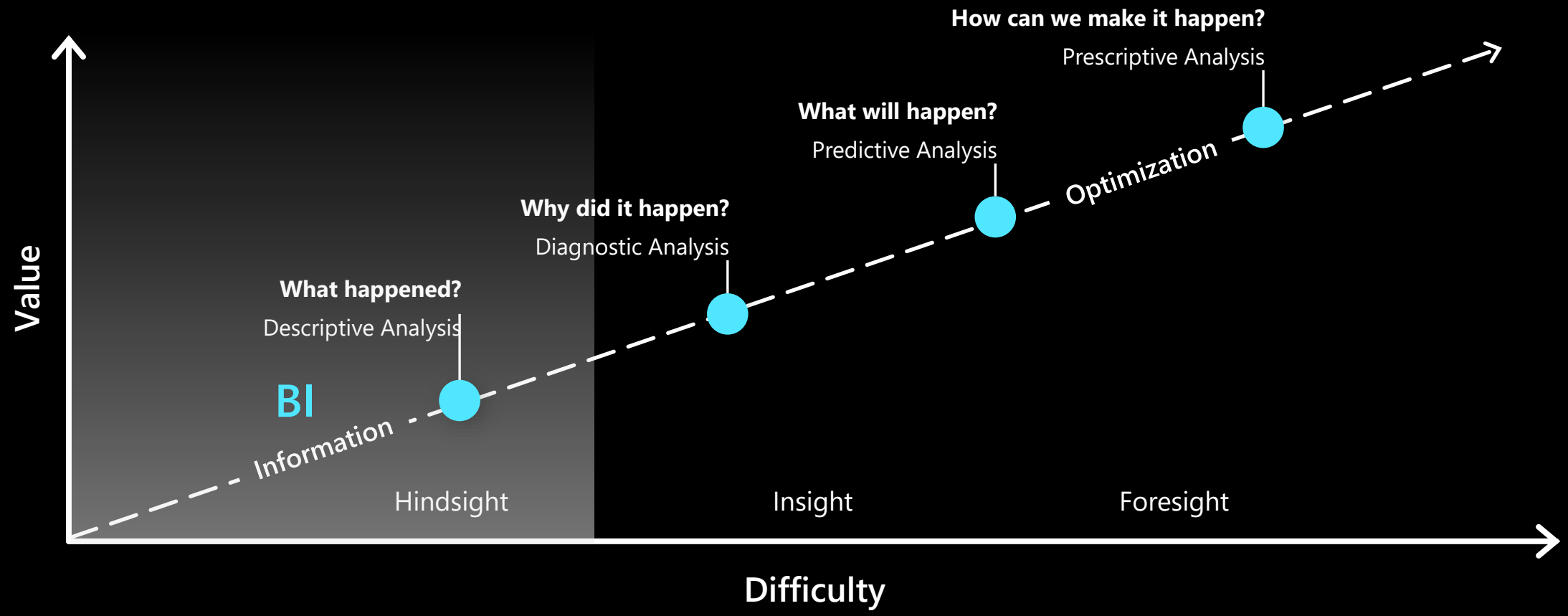
Power BI + Azure Synapse

An unmatched combination

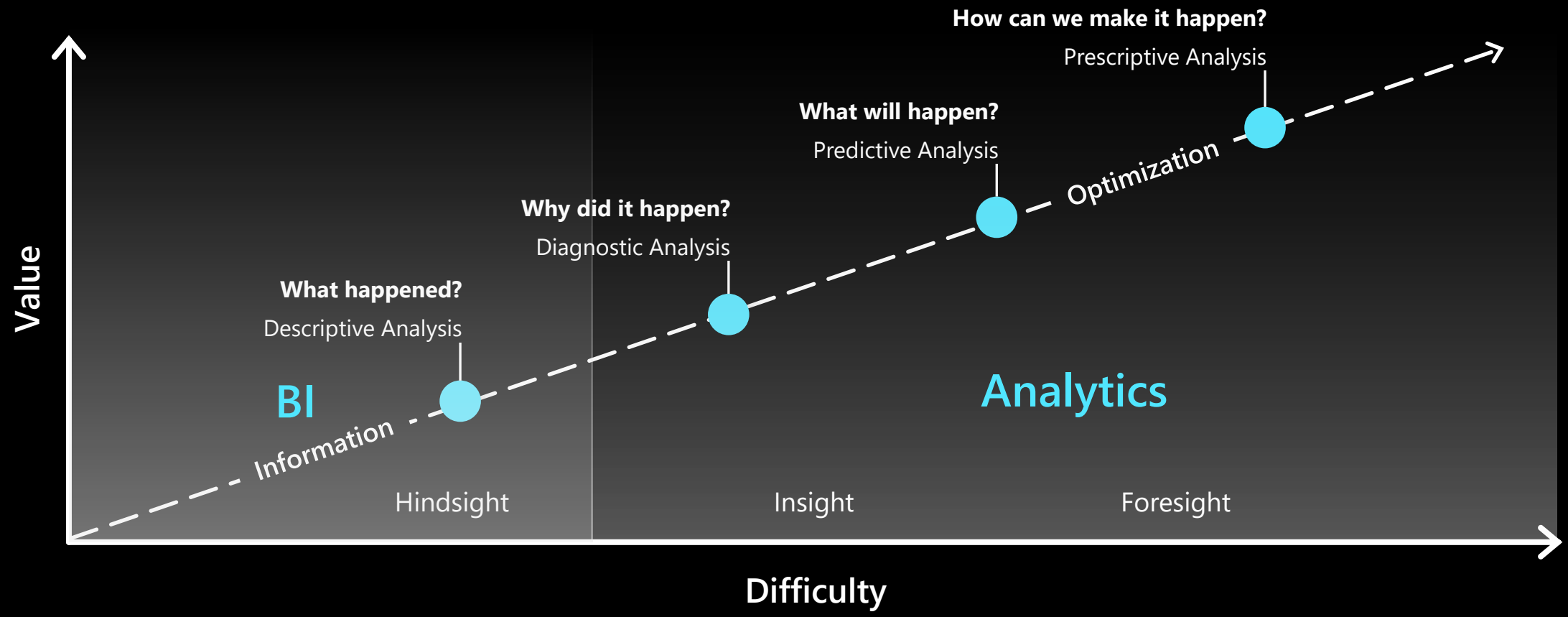
Where do you find yourself on the curve?

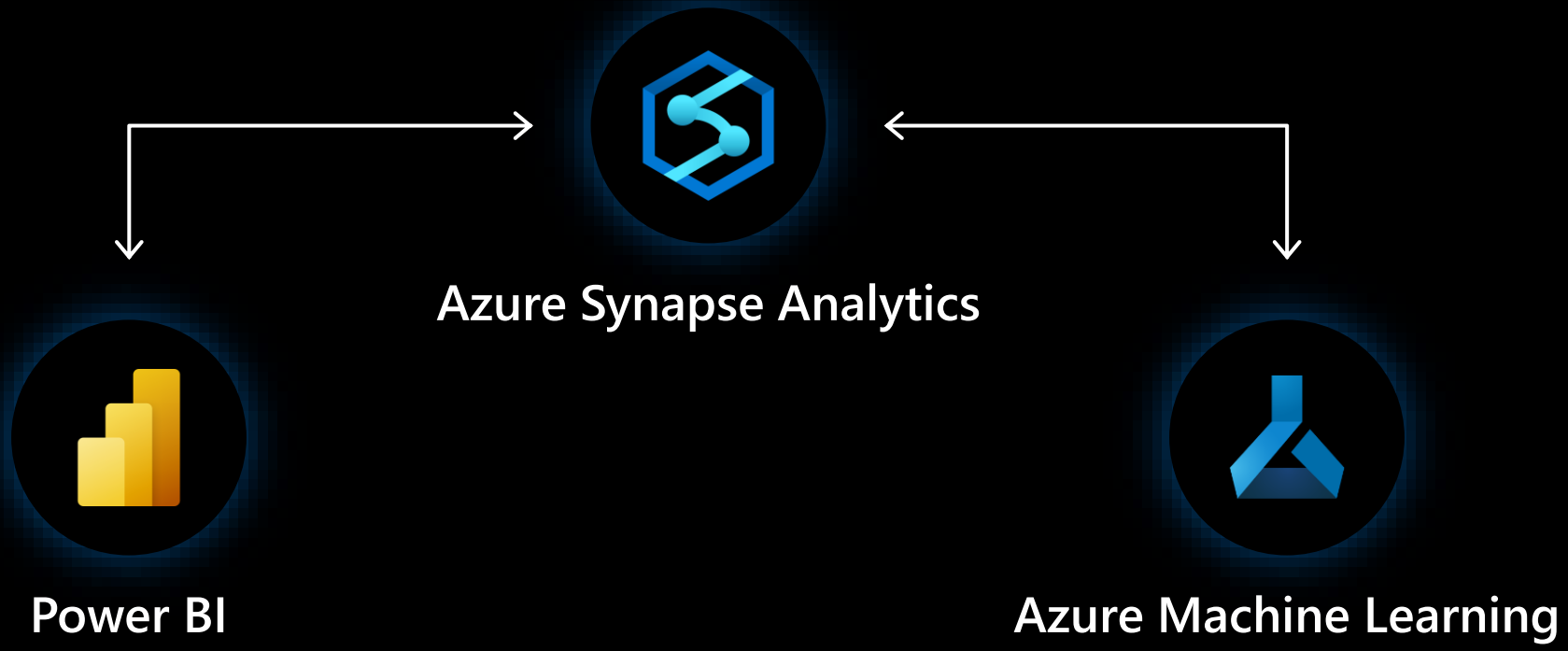


Where do you find yourself on the curve?



BI + Analytics unlock the door to AI, machine learning, and real-time insights





Unified experience to enrich data

Automated ML for rapid development

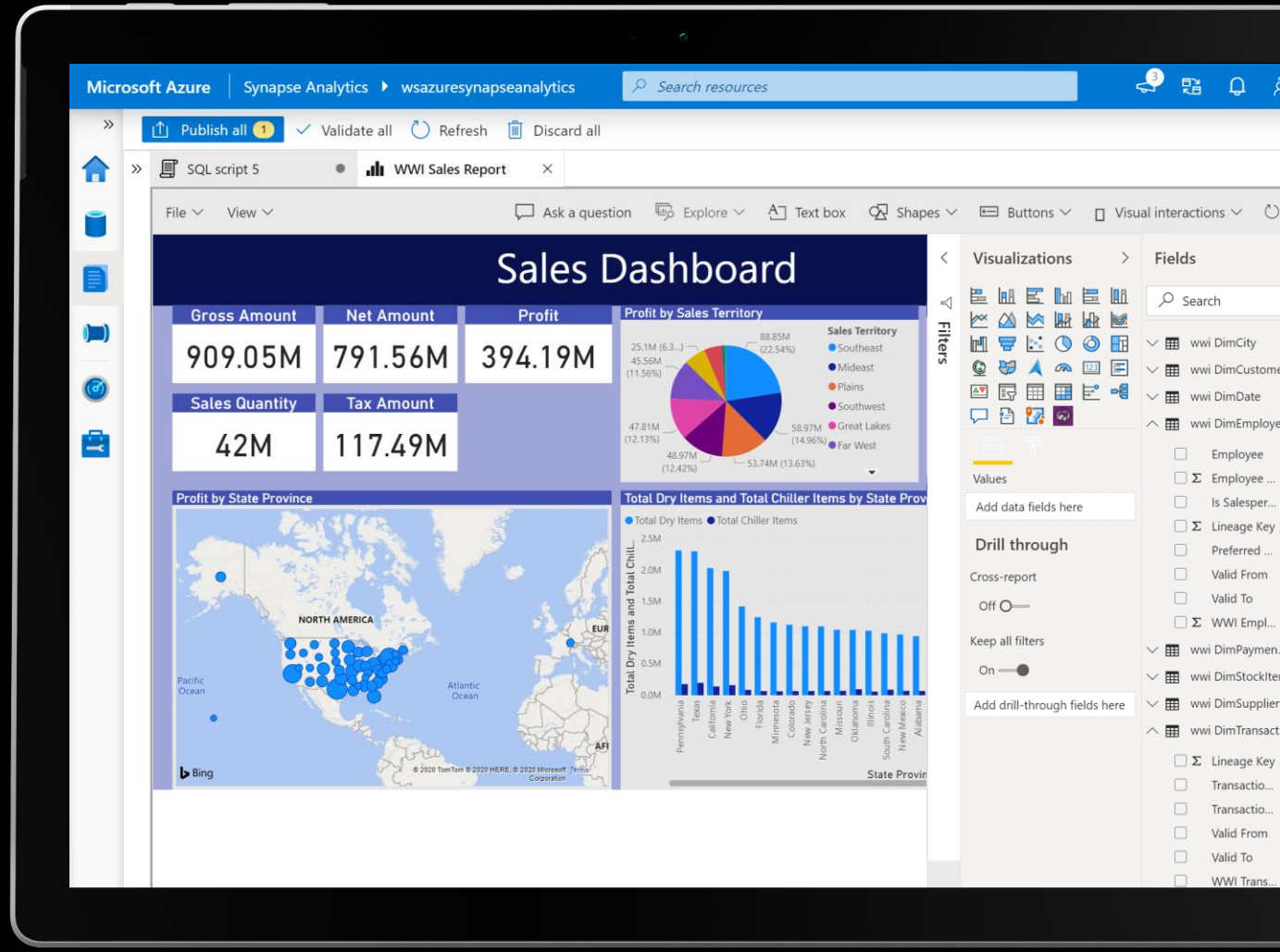
Seamless collaboration

Power BI integration

Build dashboard in Synapse Studio

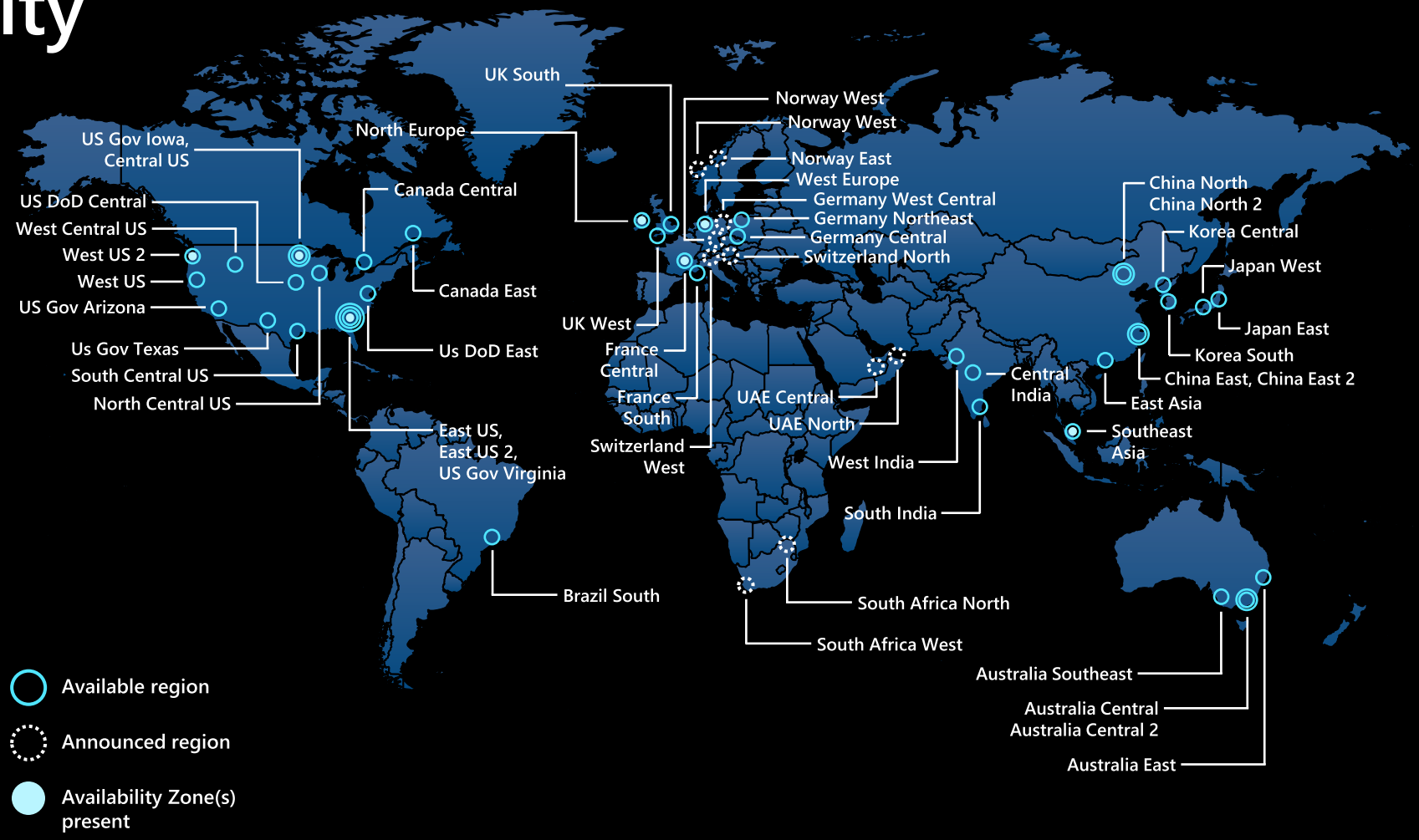
Code-free experience for development rich visualizations

One-click publishing to for secure consumption across the enterprise



Azure Synapse regional availability

Australia Southeast	Korea Central
Australia East	North Central US
Brazil South	North Europe
Canada Central	South Africa North
Canada East	South Central US
Central India	Southeast Asia
Central US	Switzerland North
East Asia	UK West
East US	UK South
East US 2	West Central US
France Central	West Europe
Germany West Central	West US
Japan East	West US 2
Japan West	



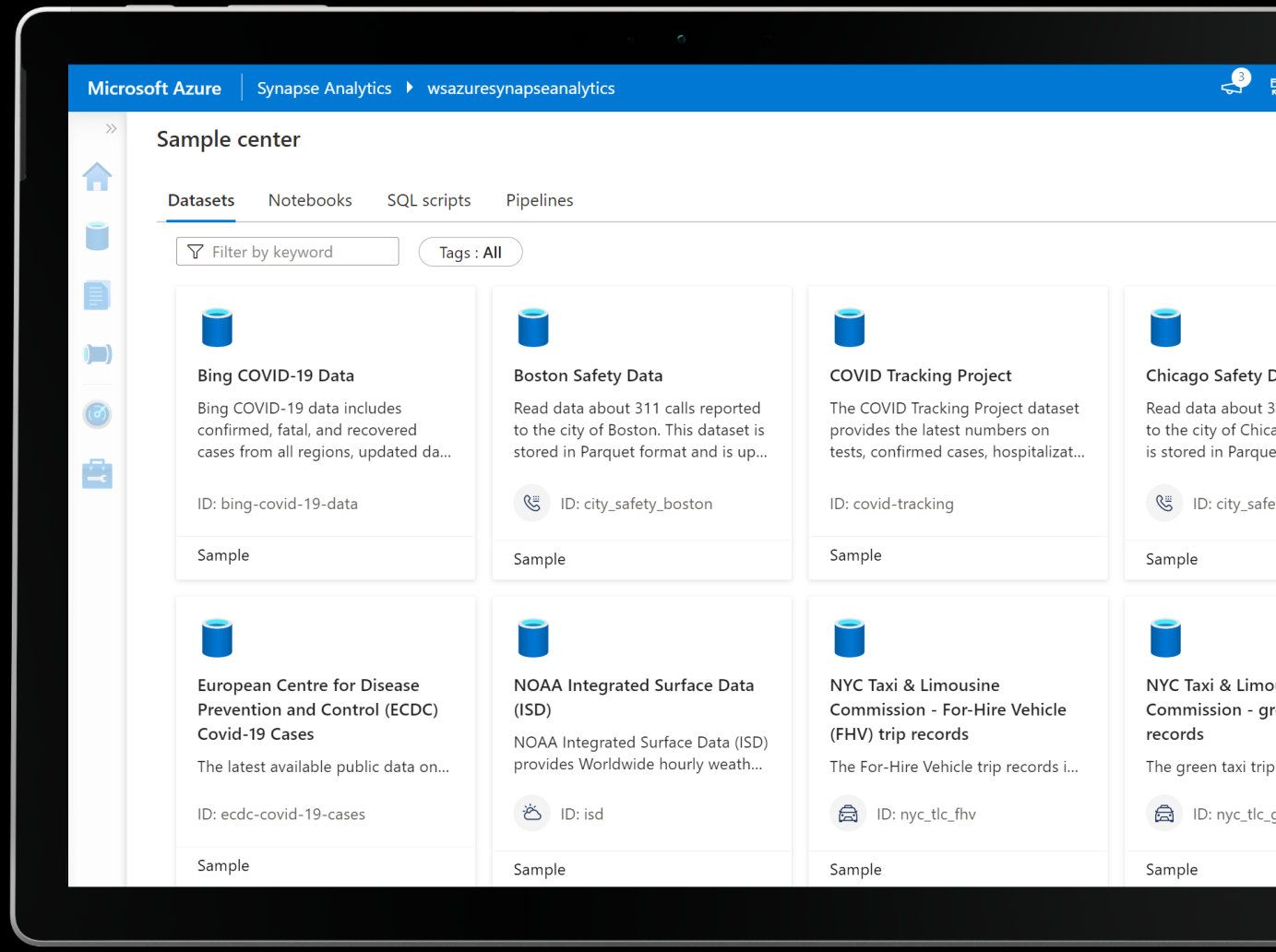
Knowledge Center

Accelerate time to solution

Azure Open Data sets

Pre-built samples to accelerate development

- SQL Scripts
- Notebooks
- Data Pipelines



Microsoft

Break

Microsoft

Hands-on lab

Build an end-to-end analytics solution with Azure Synapse & Power BI

Microsoft

Azure Immersion Workshop: Analytics | 2021-03-16 | Denmark

Dear Steen Kristensen

Your **Azure Immersion Workshop: Analytics | 2021-03-16 | Denmark** On demand lab is ready.
You have 4Hrs, 0Mins to try out the lab before it expires.

Here is the Azure login info and lab guide details:

On Demand Lab: Azure Immersion Workshop: Analytics | 2021-03-16 | Denmark
Username: [redacted]@msazurelabs.onmicrosoft.com
Password: [redacted]
Lab Guide: https://github.com/solliancenet/azure-synapse-analytics-day

Please use the below details for future use in your labs:

lab-ti-330680 :

Name	Value
uniqueId	330680
storage Account Name	asadatalake330680
labvm Admin Username	demouser
labvm Admin Password	Password.1!!
labvm DNS Name	labvm-nt3ddkcngslwk.westeurope.cloudapp.azure.com

Access Lab Now

If you have any questions, please contact us at cloudlabs-support@spektrasystems.com

Microsoft

Steen Kjøng Kristensen

Senior Consultant, Business Intelligence
Accobat A/S

skp@accobat.com
https://www.linkedin.com/in/spauls/